

High-SNR Asymptotics of Mutual Information for Discrete Constellations with Applications to BICM

Alex Alvarado, Fredrik Brännström, Erik Agrell, and Tobias Koch

Abstract—The high-signal-to-noise ratio (SNR) asymptotic behavior of the mutual information (MI) for discrete constellations over the scalar additive white Gaussian noise channel is studied. Exact asymptotic expressions for the MI for arbitrary one-dimensional constellations and input distributions are presented in the limit as the SNR tends to infinity. Using the relationship between the MI and the minimum mean-square error (MMSE), asymptotics of the MMSE are also developed. It is shown that for any input distribution, the MI, MMSE and symbol-error probability have an asymptotic behavior proportional to a Gaussian Q-function whose argument depends only on the minimum Euclidean distance of the constellation. The developed expressions are then used to study the high-SNR behavior of the generalized mutual information (GMI) for bit-interleaved coded modulation (BICM). The long-standing conjecture that, at high SNR, Gray codes are the binary labelings that maximize the BICM-GMI, is proven. It is also shown that for any equally spaced M -ary constellation, there always exists an anti-Gray code that gives the lowest BICM-GMI at high SNR.

Index Terms—Anti-Gray code, additive white Gaussian noise channel, bit-interleaved coded modulation, discrete constellations, Gray code, minimum-mean square error, mutual information, high-SNR asymptotics.

I. INTRODUCTION

In this paper we consider the additive white Gaussian noise (AWGN) channel

$$Y = \sqrt{\rho}X + Z \quad (1)$$

where X is the transmitted symbol and Z is a Gaussian random variable, independent of X , with zero mean and unit variance. The capacity of the AWGN channel in (1) is given by [1]

$$C(\rho) = \frac{1}{2} \log(1 + \gamma) \quad (2)$$

where γ is the signal-to-noise ratio (SNR). Although inputs distributed according to the Gaussian distribution attain the capacity, they suffer from several drawbacks which prevent them from being used in practical systems. Among them, especially relevant are the unbounded support and the infinite

number of bits needed to represent signal points. In practical systems, discrete distributions with a bounded support are typically preferred.

The mutual information (MI) where the input distribution is constrained to be a probability mass function (PMF) over a discrete constellation represents the maximum rate at which information can be reliably transmitted using that particular constellation. The MI for the AWGN channel as a function of the SNR for an arbitrary input PMF involves mixtures of discrete and continuous distributions, and thus, is difficult to analyze. To overcome this problem, this MI is typically assessed asymptotically at low or high SNR.

While the low-SNR asymptotics of the MI for discrete constellations are well-understood (see [1]–[4] and references therein), to the best of our knowledge, only upper and lower bounds are known for the high-SNR behavior. For example, upper and lower bounds on the MI and/or the minimum mean-square error (MMSE) in the high-SNR regime were derived in [5]–[7]. It was argued in [6, p. 1073] that for discrete constellations maximizing the MI is equivalent to minimizing both the symbol-error probability (SEP) and the MMSE. In [8, Appendix E] two constellations with different minimum Euclidean distances (MEDs) are compared, and it is shown that, for sufficiently large SNR, the constellation with larger MED gives a higher MI. Upper and lower bounds on the MI and MMSE for multiple-antenna systems over fading channels can be found in [9]–[12].

In this paper, we study high-SNR asymptotics of the MI for discrete constellations. In particular, we consider arbitrary constellations and input distributions for the channel in (1), and find exact asymptotic expressions for the MI in the limit as the SNR tends to infinity. Using the MI-MMSE relationship established in [13] (see also [14, Ch. 2]), exact asymptotic expressions for the MMSE are also developed. We prove that for any constellation and input distribution, the MI, MMSE, and SEP have an asymptotic behavior proportional to $Q(\sqrt{\rho}d/2)$, where $Q(\cdot)$ is the Gaussian Q-function and d is the MED of the constellation. In terms of the proportionality constant, and for a uniform input distribution, these results show that the constellation that maximizes the MI is the same that minimizes both the MMSE and the SEP.

While the results in this paper are general, we use them to study bit-interleaved coded modulation (BICM) [15]–[17], which can be viewed as a pragmatic approach for coded modulation. The key element in BICM is the use of a (suboptimal) bit-wise detection rule, which was cast as a mismatched decoder in [18]. BICM is used in almost all of the current wireless communications standards, e.g., HSPA, IEEE

Research supported by the European Community's Seventh's Framework Programme (FP7/2007-2013) under grant agreement No. 271986, by the Swedish Research Council, Sweden (under grants #621-2006-4872 and #621-2011-5950) and by Ministerio de Economía of Spain (projects "DEIPRO", id. TEC2009-14504-C02-01, and "COMONSENS", id. CSD2008-00010).

A. Alvarado is with the Dept. of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom (email: alex.alvarado@ieee.org).

E. Agrell and F. Brännström are with the Dept. of Signals and Systems, Chalmers Univ. of Technology, SE-41296 Göteborg, Sweden (email: {fredrik.brannstrom, agrell}@chalmers.se).

T. Koch is with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Spain (email: koch@tsc.uc3m.es).

802.11a/g/n, and the DVB standards (DVB-T2/S2/C2).

The BICM generalized mutual information (BICM-GMI) is an achievable rate for BICM systems [18] and heavily depends on the binary labeling of the constellation. The optimality of a Gray code (GC) in terms of maximizing the BICM-GMI was conjectured in [16, Sec. III-C] for general GCs, constellations, and SNR values. This conjecture was later disproved in [19]: it was shown that for low and medium SNRs, there exist other labelings that give a higher BICM-GMI (see also [20, Ch. 3]). For further results on BICM at low SNR see [21]–[24]. On the other hand, numerical results presented in [20, Ch. 3] and [25] suggest that GCs are optimal at high SNR. The proof of the optimality of GCs in terms of BICM-GMI in the high-SNR regime remains as an open problem [20, Sec. 3.3].

In this paper, we derive an asymptotic expression for the BICM-GMI as a function of the constellation, input distribution, and binary labeling. Using this expression, we then prove the optimality of GCs at high SNR. Using the MI-MMSE relationship, an asymptotic expression for the derivative of the BICM-GMI is also developed. The obtained asymptotic expressions for the BICM-GMI and its derivative, as well as the one for the bit-error probability (BEP), are all shown to be proportional to $Q(\sqrt{\rho d}/2)$.

This paper is organized as follows. In Sec. II, the notation convention and system model are presented. The asymptotics of the MI and MMSE are presented in Sec. III and BICM is studied in Sec. IV. The conclusions are drawn in Sec. V.

II. PRELIMINARIES

A. Notation Convention

Row vectors are denoted by boldface letters $\mathbf{x} = [x_1, x_2, \dots, x_M]$ and sets are denoted by calligraphic letters \mathcal{C} . An exception is the set of real numbers, which is denoted by \mathbb{R} . The binary set is defined as $\mathcal{B} \triangleq \{0, 1\}$ and the bipolar set by $\mathcal{W} \triangleq \{-1, +1\}$. The negation of a bit b is denoted by \bar{b} . All the logarithms are natural logarithms and all the MIs are therefore given in nats. Marginal probability density functions (PDFs) and conditional PDFs are denoted by $f_Y(y)$ and $f_{Y|X}(y|x)$, respectively. Analogously, PMFs are denoted by $P_X(x)$ and $P_{X|Y}(x|y)$. Expectations are denoted by $\mathbb{E}[\cdot]$.

B. Model

We consider the discrete-time, real-valued AWGN channel in (1), where the transmitted symbols X are constrained to $X \in \mathcal{X} \triangleq \{x_1, x_2, \dots, x_{|\mathcal{X}|}\}$, and where $|\mathcal{X}| = M = 2^m$. The set of indices that enumerates all the constellation symbols in \mathcal{X} is defined as $\mathcal{I}_{\mathcal{X}} \triangleq \{1, \dots, M\}$.

We focus on one-dimensional constellations and assume, without loss of generality, that the symbols are different and ordered, i.e., $x_1 < x_2 < \dots < x_M$. Each of the symbols is transmitted with probability $p_i \triangleq P_X(x_i)$, $0 < p_i < 1$. While the transmitted symbols are fully determined by the PMF P_X , we shall use *constellation* to denote the support \mathcal{X} of the PMF and *input distribution* to denote the probabilities $[p_1, \dots, p_M]$ associated with the symbols. We assume that neither the constellation nor the input distribution depends on ρ .

The transmitted average symbol energy is finite and given by

$$E_s \triangleq \mathbb{E}[X^2] = \sum_{i=1}^M p_i x_i^2. \quad (3)$$

It follows that the SNR γ in (2) is $\gamma = \rho E_s$.

An M -ary pulse-amplitude modulation (MPAM) constellation having M equally spaced symbols (separated by 2Δ) is denoted by $\mathcal{E} \triangleq \{x_i = -(M - 2i + 1)\Delta : i = 1, \dots, M\}$. A uniform distribution over X is denoted by P_X^u , i.e., $p_i = 1/M \forall i$. A uniform input distribution with $\mathcal{X} = \mathcal{E}$ is denoted by P_X^u , where in this case $\Delta^2 = 3E_s/(M^2 - 1)$.

The Gaussian Q-function is defined as

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}\xi^2} d\xi \quad (4)$$

the entropy of the random variable X as

$$H_{P_X} \triangleq -\mathbb{E}[\log(P_X(X))] \quad (5)$$

the MI between X and Y as

$$I_{P_X}(\rho) \triangleq \mathbb{E}[\log(f_{Y|X}(Y|X)/f_Y(Y))] \quad (6)$$

and the MMSE as

$$M_{P_X}(\rho) \triangleq \mathbb{E}[(X - \hat{X}^{\text{ME}}(Y))^2] \quad (7)$$

where $\hat{X}^{\text{ME}}(y) \triangleq \mathbb{E}[X|Y = y]$ is the conditional (posterior) mean estimator.

We also define the SEP as

$$S_{P_X}(\rho) \triangleq \Pr\{\hat{X}^{\text{MAP}}(Y) \neq X\} \quad (8)$$

where X is the transmitted symbol and

$$\hat{X}^{\text{MAP}}(y) \triangleq \underset{x \in \mathcal{X}}{\operatorname{argmax}} P_{X|Y}(x|y) \quad (9)$$

is the decision made by a maximum *a posteriori* probability (MAP) symbol demapper.

C. Discrete Constellations

We define the set of *symbol differences* corresponding to $x_i \in \mathcal{X}$ as

$$\mathcal{D}_{\mathcal{X}}^{(i)} \triangleq \{x_i - x : x \in \mathcal{X}\} \quad (10)$$

and the set of all symbol differences in \mathcal{X} as

$$\mathcal{D}_{\mathcal{X}} \triangleq \bigcup_{i \in \mathcal{I}_{\mathcal{X}}} \mathcal{D}_{\mathcal{X}}^{(i)}. \quad (11)$$

The MED of the constellation is defined as

$$d \triangleq \min_{\delta \in \mathcal{D}_{\mathcal{X}} \setminus \{0\}} |\delta| \quad (12)$$

and the maximum Euclidean distance (ED) of the constellation as

$$\hat{d} \triangleq \max_{\delta \in \mathcal{D}_{\mathcal{X}}} |\delta|. \quad (13)$$

We further define the counting function

$$A_{\mathcal{X}}^{(i)}(\delta) \triangleq \begin{cases} 1, & \text{if } \delta \in \mathcal{D}_{\mathcal{X}}^{(i)} \\ 0, & \text{if } \delta \notin \mathcal{D}_{\mathcal{X}}^{(i)} \end{cases}. \quad (14)$$

Since 0 is always an element of $\mathcal{D}_{\mathcal{X}}^{(i)}$, we have $A_{\mathcal{X}}^{(i)}(0) = 1 \forall i$.

Analogous to $A_{\mathcal{X}}^{(i)}(\delta)$, we define $B_{P_X}^{(i)}(\delta)$ as

$$B_{P_X}^{(i)}(\delta) \triangleq \begin{cases} \sqrt{p_j/p_i}, & \text{if } \exists x_j \in \mathcal{X} : x_i - x_j = \delta \\ 0, & \text{otherwise} \end{cases}. \quad (15)$$

Clearly $B_{P_X}^{(i)}(0) = 1, i \in \mathcal{I}_{\mathcal{X}}$.

We define $A_{\mathcal{X}}$ as twice the number of pairs of constellation points at MED, i.e.,

$$A_{\mathcal{X}} \triangleq \sum_{i \in \mathcal{I}_{\mathcal{X}}} \sum_{w \in \mathcal{W}} A_{\mathcal{X}}^{(i)}(wd). \quad (16)$$

By using the fact that for any real-valued constellation there are at least one and at most $M-1$ pairs of constellation points at MED, we obtain the bound

$$2 \leq A_{\mathcal{X}} \leq 2(M-1). \quad (17)$$

The upper bound is achieved by an MPAM constellation

$$A_{\mathcal{E}} = 2(M-1). \quad (18)$$

Finally, for a given P_X , we define the constants

$$N_{P_X} \triangleq \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \sum_{w \in \mathcal{W}} B_{P_X}^{(i)}(wd) \quad (19)$$

and

$$D_{P_X} \triangleq \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \sum_{w \in \mathcal{W}} A_{\mathcal{X}}^{(i)}(wd). \quad (20)$$

For a uniform input distribution $P_X = P_X^u$ and $B_{P_X}^{(i)}(\delta) = A_{\mathcal{X}}^{(i)}(\delta)$, so

$$N_{P_X^u} = D_{P_X^u} = \frac{A_{\mathcal{X}}}{M}. \quad (21)$$

Example 1: Consider a nonequally spaced 4-ary constellation with $x_1 = -5$, $x_2 = -3$, $x_3 = 3$, and $x_4 = 5$, and the input distribution $p_i = i/10$ with $i = 1, 2, 3, 4$. In this case,

$$\begin{aligned} \mathcal{D}_{\mathcal{X}}^{(1)} &= \{0, -2, -8, -10\}, \\ \mathcal{D}_{\mathcal{X}}^{(2)} &= \{2, 0, -6, -8\}, \\ \mathcal{D}_{\mathcal{X}}^{(3)} &= \{8, 6, 0, -2\}, \\ \mathcal{D}_{\mathcal{X}}^{(4)} &= \{10, 8, 2, 0\} \end{aligned}$$

and $\mathcal{D}_{\mathcal{X}} = \{0, \pm 2, \pm 6, \pm 8, \pm 10\}$. The MED of the constellation in (12) is $d = 2$, E_s in (3) is $E_s = 17$, $A_{\mathcal{X}}$ in (16) is $A_{\mathcal{X}} = 4$ (two pairs of constellation points at MED), N_{P_X} in (19) is $N_{P_X} = 2\sqrt{p_1 p_2} + 2\sqrt{p_3 p_4} \approx 0.98$, and D_{P_X} in (20) is $D_{P_X} = p_2 + p_1 + p_4 + p_3 = 1$.

III. HIGH-SNR ASYMPTOTICS

There exists a fundamental relationship between the MI and the MMSE for AWGN channels [13] (see also [14, Ch. 2]):

$$\frac{d}{d\rho} \mathcal{I}_{P_X}(\rho) = \frac{1}{2} \mathcal{M}_{P_X}(\rho). \quad (22)$$

Exploiting this MI-MMSE relation, bounds on the MI can be used to derive bounds on the MMSE and *vice versa*.

Upper and lower bounds on the MI and MMSE for discrete constellations at high SNR can be found in e.g., [5]–[7], [9]–[12]. While these bounds describe the correct asymptotic behavior, they are, in general, not tight in the sense that the ratio between them does not tend to one as $\rho \rightarrow \infty$. In what follows, we present exact asymptotic expressions for the MI and MMSE for an arbitrary P_X .

A. Asymptotics of the MI, MMSE, and SEP

For any given input distribution P_X , the MI tends to \mathcal{H}_{P_X} as ρ tends to infinity. In the following we study how fast the MI converges towards its maximum \mathcal{H}_{P_X} by analyzing the difference $\mathcal{H}_{P_X} - \mathcal{I}_{P_X}(\rho)$.¹ Theorem 1 is the main result of this paper and characterizes the high-SNR behavior of $\mathcal{H}_{P_X} - \mathcal{I}_{P_X}(\rho)$.

Theorem 1: For any P_X

$$\lim_{\rho \rightarrow \infty} \frac{\mathcal{H}_{P_X} - \mathcal{I}_{P_X}(\rho)}{\mathcal{Q}(\sqrt{\rho}d/2)} = \pi N_{P_X} \quad (23)$$

where N_{P_X} is given in (19).

Proof: The proof is given in Appendix A. ■

Both numerator and denominator on the left-hand side (l.h.s.) of (23) tend to zero as ρ tends to infinity. It follows from L'Hôpital's rule that the limiting ratio in (23) is equal to the limiting ratio of the numerator's and denominator's derivatives. By the MI-MMSE relation in (22), Theorem 1 thus directly gives an asymptotic expression for the MMSE.

Theorem 2: For any P_X

$$\lim_{\rho \rightarrow \infty} \frac{\mathcal{M}_{P_X}(\rho)}{\mathcal{Q}(\sqrt{\rho}d/2)} = \frac{d^2}{4} \pi N_{P_X} \quad (24)$$

where N_{P_X} is given by (19).

Proof: The proof follows by applying L'Hôpital's rule to the l.h.s. of (23) together with (22), by recognizing the derivative of $\mathcal{Q}(\sqrt{\rho}d/2)$ in (23) as $-G(\sqrt{\rho}d/2) d^2/8$ with $G(x)$ defined in (83), and by using (84). ■

The asymptotic expression for the SEP is stated in the following theorem.

Theorem 3: For any P_X

$$\lim_{\rho \rightarrow \infty} \frac{S_{P_X}(\rho)}{\mathcal{Q}(\sqrt{\rho}d/2)} = D_{P_X} \quad (25)$$

where D_{P_X} is given in (20).

Proof: The proof is given in Appendix B. ■

Theorems 1–3 reveal that, at high SNR, the MI, MMSE, and SEP behave as

$$\mathcal{I}_{P_X}(\rho) \approx \mathcal{H}_{P_X} - \pi N_{P_X} \mathcal{Q}\left(\frac{\sqrt{\rho}d}{2}\right), \quad (26)$$

$$\mathcal{M}_{P_X}(\rho) \approx \frac{d^2}{4} \pi N_{P_X} \mathcal{Q}\left(\frac{\sqrt{\rho}d}{2}\right), \quad (27)$$

$$S_{P_X}(\rho) \approx D_{P_X} \mathcal{Q}\left(\frac{\sqrt{\rho}d}{2}\right). \quad (28)$$

The results in (26)–(28) show that for any input distribution, the MI, MMSE, and SEP have the same high-SNR behavior,

¹The quantity $\mathcal{H}_{P_X} - \mathcal{I}_{P_X}(\rho)$ corresponds to the conditional entropy of X given Y .

i.e., they are all proportional to a Gaussian Q-function², where the proportionality constants depend on the input distribution.

Remark 1: While the results presented in this section hold for any one-dimensional constellation, they directly generalize to multidimensional constellations that are constructed as *ordered direct products* [23, eq. (1)] of one-dimensional constellations. For example, the results directly generalize to rectangular quadrature amplitude modulation constellations.

B. Discussion and Examples

For a uniform input distribution ($P_X = P_X^u$), Theorems 1–3 particularize to the following result.

Corollary 1: For any \mathcal{X} with a uniform input distribution

$$\lim_{\rho \rightarrow \infty} \frac{\log M - I_{P_X^u}(\rho)}{Q(\sqrt{\rho d}/2)} = \pi \frac{A_{\mathcal{X}}}{M}, \quad (29)$$

$$\lim_{\rho \rightarrow \infty} \frac{M_{P_X^u}(\rho)}{Q(\sqrt{\rho d}/2)} = \frac{d^2}{4} \pi \frac{A_{\mathcal{X}}}{M}, \quad (30)$$

$$\lim_{\rho \rightarrow \infty} \frac{S_{P_X^u}(\rho)}{Q(\sqrt{\rho d}/2)} = \frac{A_{\mathcal{X}}}{M} \quad (31)$$

where $A_{\mathcal{X}}$ is given in (16).

Proof: Follow directly from (21). ■

The expression (31) corresponds to the well-known high-SNR approximation for the SEP [26, eq. (2.3–29)]. Moreover, Corollary 1 shows that for a uniform input distribution, the MI, the MMSE, and the SEP for discrete constellations in the high-SNR regime are functions of the MED of the constellation and by the number of pairs of constellation points at MED only.

For MPAM and a uniform input distribution ($P_X = P_X^u$), Corollary 1 particularizes to (see (18))

$$\lim_{\rho \rightarrow \infty} \frac{\log M - I_{P_X^u}(\rho)}{Q(\sqrt{\rho d}/2)} = \pi \lim_{\rho \rightarrow \infty} \frac{S_{P_X^u}(\rho)}{Q(\sqrt{\rho d}/2)} \quad (32)$$

$$= \pi \frac{2(M-1)}{M} \quad (33)$$

and

$$\lim_{\rho \rightarrow \infty} \frac{M_{P_X^u}(\rho)}{Q(\sqrt{\rho d}/2)} = \pi \frac{6E_s}{M(M+1)}. \quad (34)$$

In Table I, we show a summary of the results obtained in this section, including (32)–(34).

Example 2: In Fig. 1, we show the conditional entropy $\log M - I_{P_X^u}(\rho)$ for 4PAM and 16PAM with uniform input distributions³ together with the asymptotic expression in (33). We also show the lower and upper bounds derived in [6, eq. (34)–(35)] and [12, eq. (17)–(19)]. Observe that (33) approximates $\log M - I_{P_X^u}(\rho)$ accurately for a large range of SNR. In Fig. 2, analogous results for the MMSE are presented, where the bounds derived in [6, eq. (30)–(31)] and [12, eq. (13)–(15)] are also included. Again our asymptotic expression (34) approximates the MMSE accurately for a large range of SNR.

²Disregarding the “offset” H_{P_X} in (26).

³Calculated numerically using Gauss–Hermite quadratures with 300 quadrature points [25, Sec. III].

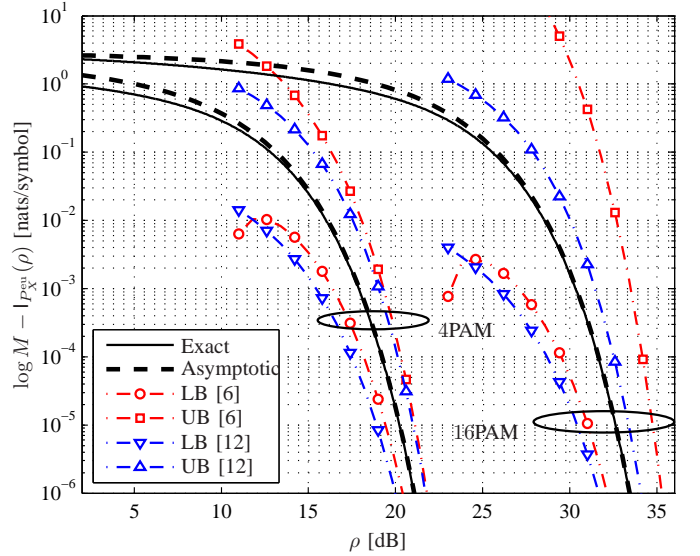


Fig. 1. $\log M - I_{P_X^u}(\rho)$ for 4PAM and 16PAM (solid lines) constellations (normalized to $E_s = 1$) and the asymptotic expression in (33) (thick dashed lines). The lower and upper bounds [6, eq. (34)–(35)] and [12, eq. (17)–(19)] are also shown.

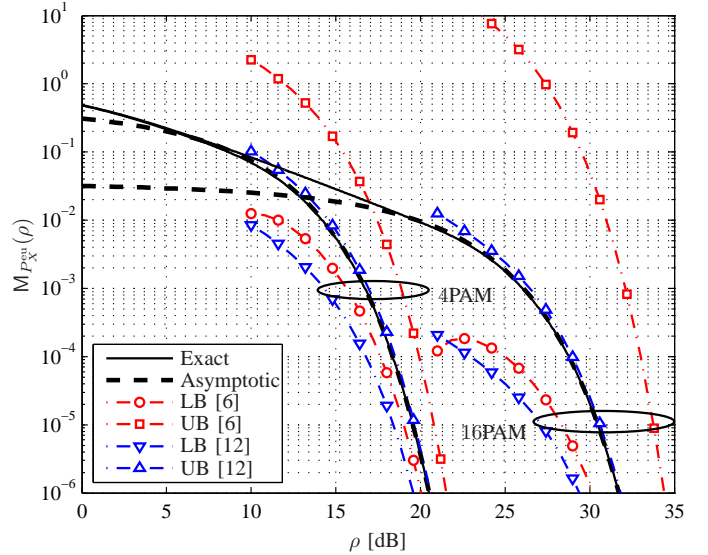


Fig. 2. $M_{P_X^u}(\rho)$ for 4PAM and 16PAM (solid lines) constellations (normalized to $E_s = 1$) and the asymptotic expression in (34) (thick dashed lines). The lower and upper bounds [6, eq. (30)–(31)] and [12, eq. (13)–(15)] are also shown.

Remark 2: It follows from Corollary 1 that the constellation that maximizes the MI (or equivalently, the constellation that minimizes the MMSE and the SEP) at high SNR is the constellation that first maximizes the MED and then minimizes $A_{\mathcal{X}}$. For one-dimensional constellations, the MED is maximized by an MPAM constellation ($\mathcal{X} = \mathcal{E}$).

We conclude this section by noting that if Theorems 1 and 2 are combined, we obtain

$$\lim_{\rho \rightarrow \infty} \frac{M_{P_X}(\rho)}{H_{P_X} - I_{P_X}(\rho)} = \frac{d^2}{4}. \quad (35)$$

Thus, for any P_X , the limiting ratio between the MMSE and the conditional entropy is constant, regardless of the input

TABLE I
SUMMARY OF ASYMPTOTICS OF MI, MMSE, AND SEP.

Input Distribution	P_X	P_X^u	P_X^{eu}
$\lim_{\rho \rightarrow \infty} \frac{H_{P_X} - I_{P_X}(\rho)}{Q(\sqrt{\rho d}/2)}$	πN_{P_X}	$\pi \frac{A_X}{M}$	$\pi \frac{2(M-1)}{M}$
$\lim_{\rho \rightarrow \infty} \frac{M_{P_X}(\rho)}{Q(\sqrt{\rho d}/2)}$	$\frac{d^2}{4} \pi N_{P_X}$	$\frac{d^2}{4} \pi \frac{A_X}{M}$	$\pi \frac{6E_s}{M(M+1)}$
$\lim_{\rho \rightarrow \infty} \frac{S_{P_X}(\rho)}{Q(\sqrt{\rho d}/2)}$	D_{P_X}	$\frac{A_X}{M}$	$\frac{2(M-1)}{M}$

distribution. Moreover, using Theorems 1 and 3, we obtain

$$\lim_{\rho \rightarrow \infty} \frac{H_{P_X} - I_{P_X}(\rho)}{S_{P_X}(\rho)} = \pi \frac{N_{P_X}}{D_{P_X}} \quad (36)$$

which for a uniform input distribution particularizes to

$$\lim_{\rho \rightarrow \infty} \frac{\log M - I_{P_X^u}(\rho)}{S_{P_X^u}(\rho)} = \pi. \quad (37)$$

Thus, for any P_X , the limiting ratio between the conditional entropy and the SEP is constant, which for a uniform input distribution equals π .

IV. APPLICATION: BINARY LABELINGS FOR BIT-INTERLEAVED CODED MODULATION

We next study the high-SNR behavior of BICM systems [15]–[17]. In such systems (see Fig. 3) the bits are mapped to constellation symbols using a mapper $\Phi: \mathcal{B}^m \rightarrow \mathcal{X}$. Using the results in Sec. III, we will find an asymptotic expression for the BICM-GMI and we will study the relationship between the BICM-GMI and the BEP. We will also prove the asymptotic optimality of GCs in terms of BICM-GMI for one-dimensional constellations with uniform input distributions.

A. BICM Model

A binary labeling for a constellation is defined by the vector $\mathbf{l} = [l_1, l_2, \dots, l_M]$ where $l_i \in \{0, 1, \dots, M-1\}$ is the integer representation of the i th length- m binary label $\mathbf{q}_i = [q_{i,1}, \dots, q_{i,m}] \in \mathcal{B}^m$ associated with the symbol x_i , with $q_{i,1}$ being the most significant bit. The labeling defines $2m$ subconstellations $\mathcal{X}_{k,b} \subset \mathcal{X}$ for $k = 1, \dots, m$ and $b \in \mathcal{B}$, i.e., $\mathcal{X}_{k,b} \triangleq \{x_i \in \mathcal{X} : q_{i,k} = b\}$ with $|\mathcal{X}_{k,b}| = M/2$. We define $\mathcal{I}_{\mathcal{X}_{k,b}} \subset \{1, \dots, M\}$ as the indices of the symbols in $\mathcal{X}_{k,b}$.

In BICM, the coded bits $\mathbf{Q} = [Q_1, Q_2, \dots, Q_m]$ at the input of the mapper (see Fig. 3) are assumed to be independent but possibly nonuniformly distributed. Therefore, the vector of bit probabilities $[P_{Q_1}(0), P_{Q_2}(0), \dots, P_{Q_m}(0)]$ induces a symbol input distribution P_X via the labeling as [23, eq. (31)] [27, eq. (8)]

$$P_X(x_i) = p_i = \prod_{k=1}^m P_{Q_k}(q_{i,k}). \quad (38)$$

Using (38), we obtain the conditional probabilities

$$P_{X|Q_k}(x|b) = \begin{cases} \frac{P_X(x)}{P_{Q_k}(b)}, & \text{if } x \in \mathcal{X}_{k,b} \\ 0, & \text{if } x \notin \mathcal{X}_{k,b} \end{cases} \quad (39)$$

for $k = 1, \dots, m$ and $b \in \mathcal{B}$. According to (39), each of the $2m$ conditional input distributions $[P_{X|Q_k}(x_1|b), \dots, P_{X|Q_k}(x_M|b)]$ has $M/2$ non-zero probabilities, which specify which of the $M/2$ symbols in \mathcal{X} are included in $\mathcal{X}_{k,b}$.

For uniformly distributed bits, i.e., $P_{Q_k}(b) = 1/2$, it follows that the symbol distribution is also uniform, i.e., $P_X = P_X^u$, and thus,

$$P_{X|Q_k}(x_i|b) = \begin{cases} \frac{2}{M}, & \text{if } x_i \in \mathcal{X}_{k,b} \\ 0, & \text{if } x_i \notin \mathcal{X}_{k,b} \end{cases}. \quad (40)$$

We shall use $X_{k,b}$ to denote a random variable with support $\mathcal{X}_{k,b}$ and probabilities $P_{X|Q_k}(x|b)$ for $x \in \mathcal{X}_{k,b}$ in (39). The corresponding PMF is denoted by $P_{X_{k,b}}$ and the PMF for the uniform case in (40) is denoted by $P_{X_{k,b}}^u$.

In what follows, we will apply the results of Sec. III to BICM. To this end, we will often replace \mathcal{X} and P_X in Sec. III by $\mathcal{X}_{k,b}$ and $P_{X_{k,b}}$, respectively. Note, however, that d and \hat{d} , as defined respectively in (12) and (13), still denote the MED and maximum ED of the constellation \mathcal{X} . We will not consider the MED or maximum ED for subconstellations. This implies that it is possible that no pairs of constellation points in $\mathcal{X}_{k,b}$ are at MED. Consequently, the bounds in (17) become

$$0 \leq A_{\mathcal{X}_{k,b}} \leq 2(M/2 - 1). \quad (41)$$

Moreover, when subconstellations are considered, N_{P_X} in (19) becomes

$$N_{P_{X_{k,b}}} = \sum_{i \in \mathcal{I}_{\mathcal{X}_{k,b}}} P_{X_{k,b}}(x_i) \sum_{w \in \mathcal{W}} B_{P_{X_{k,b}}}^{(i)}(wd) \quad (42)$$

and for a uniform input distribution,

$$N_{P_{X_{k,b}}^u} = \frac{A_{\mathcal{X}_{k,b}}}{M/2}. \quad (43)$$

Example 3: In Fig. 4, we show the $2m = 6$ subconstellations for an 8PAM constellation labeled by the binary-reflected Gray code (BRGC) $\mathbf{l} = [0, 1, 3, 2, 6, 7, 5, 4]$ [28]–[30], as well as the corresponding values of $\mathcal{I}_{\mathcal{X}_{k,b}}$ and $A_{\mathcal{X}_{k,b}}$.

B. Binary Labelings for BICM

The *natural binary code (NBC)* [23, Sec. II-B] is defined as the binary labeling \mathbf{l} where $l_i = i - 1$, for $i = 1, 2, \dots, M$. The NBC is an important labeling for BICM because it is the unique optimal labeling for BICM in the low-SNR regime for $\mathcal{X} = \mathcal{E}$ and uniformly distributed bits [23, Theorem 14]. These results were extended to arbitrary distributions in [24]. A labeling \mathbf{l} is said to be a GC if for all i, j such that $|x_i - x_j| = d$, the binary labels \mathbf{q}_i and \mathbf{q}_j are at Hamming distance one. One of the most popular GCs is the BRGC [28]–[30], which we showed in Example 3 for $M = 8$.

To characterize binary labelings we define

$$C_{\mathcal{X}, \mathbf{l}} \triangleq \sum_{k=1}^m \sum_{b \in \mathcal{B}} \sum_{i \in \mathcal{I}_{\mathcal{X}_{k,b}}} \sum_{w \in \mathcal{W}} A_{\mathcal{X}_{k,b}}^{(i)}(wd) \quad (44)$$

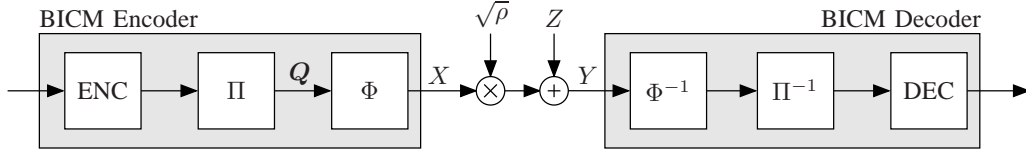


Fig. 3. A BICM scheme: The BICM encoder is formed by a serial concatenation of a binary encoder (ENC), a bit-level interleaver (Π), and a memoryless mapper (Φ). The BICM decoder is based on a demapper (Φ^{-1}) that computes logarithmic likelihood ratios, a de-interleaver (Π^{-1}), and a channel decoder (DEC).

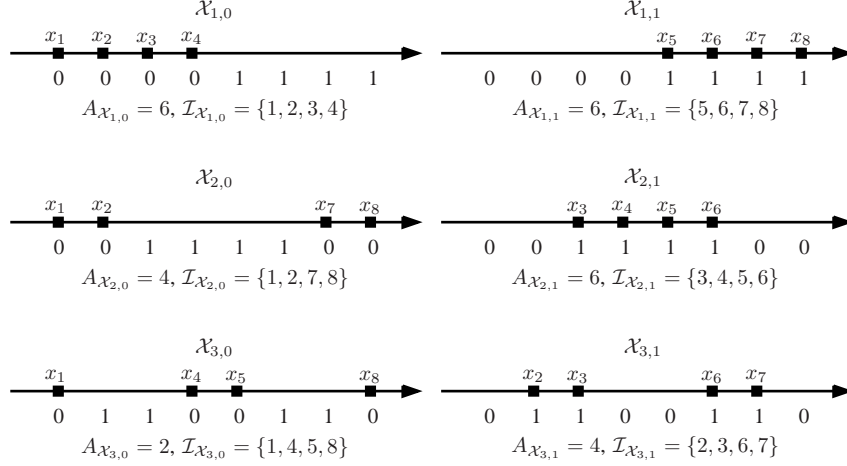


Fig. 4. Subconstellations $\mathcal{X}_{k,b}$ for 8PSK labeled by the BRGC $\mathbf{l} = [0, 1, 3, 2, 6, 7, 5, 4]$, $A_{\mathcal{X}} = C_{\mathcal{X},\mathbf{l}} = 14$. The values of $A_{\mathcal{X}_{k,b}}$ and $\mathcal{I}_{\mathcal{X}_{k,b}}$ are also shown.

and

$$E_{P_X, \mathbf{l}} \triangleq \frac{1}{m} \sum_{k=1}^m \sum_{b \in \mathcal{B}} \sum_{i \in \mathcal{I}_{\mathcal{X}_{k,b}}} p_i \sum_{w \in \mathcal{W}} A_{\mathcal{X}_{k,b}}^{(i)}(wd) \quad (45)$$

where for a uniform input distribution ($p_i = 1/M$)

$$E_{P_X, \mathbf{l}}^u = \frac{C_{\mathcal{X}, \mathbf{l}}}{mM}. \quad (46)$$

For a given subconstellation $\mathcal{X}_{k,b}$, the two inner sums in (44) consider all the constellation points in the subconstellation $\mathcal{X}_{k,b}$ at MED from $x_i \in \mathcal{X}_{k,b}$. Thus, the quantity $C_{\mathcal{X}, \mathbf{l}}$ corresponds to twice the total number of *different* bits between the labels of constellation symbol pairs at MED. Using this interpretation, it follows that (44) can also be expressed as

$$C_{\mathcal{X}, \mathbf{l}} = \sum_{k=1}^m (A_{\mathcal{X}} - A_{\mathcal{X}_{k,0}} - A_{\mathcal{X}_{k,1}}) \quad (47)$$

where $A_{\mathcal{X}} - A_{\mathcal{X}_{k,0}} - A_{\mathcal{X}_{k,1}}$ corresponds to twice the number of pairs of constellation points at MED with different labeling at bit position k . For example, for the constellation and labeling in Fig. 4, $C_{\mathcal{X}, \mathbf{l}} = 14 = A_{\mathcal{X}}$.

While $A_{\mathcal{X}}$ in (16) depends only on the geometry of the constellation, $C_{\mathcal{X}, \mathbf{l}}$ in (44) depends on both the geometry of the constellation and the labeling. By noting that any pair of constellation points at MED will differ in at least one bit, we

obtain that for any \mathcal{X} and \mathbf{l}

$$C_{\mathcal{X}, \mathbf{l}} \geq A_{\mathcal{X}}. \quad (48)$$

For $\mathcal{X} = \mathcal{E}$ and the NBC, $C_{\mathcal{X}, \mathbf{l}}$ can be expressed as

$$\begin{aligned} C_{\mathcal{E}, \mathbf{l}_{\text{NBC}}} &= 2 \sum_{k=1}^m (2^k - 1) \\ &= 2(2M - m - 2) \end{aligned} \quad (49)$$

which is obtained by noting that, for each k , there are $2^k - 1$ symbols satisfying $q_{i,k} \neq q_{i+1,k}$, for $i = 1, 2, \dots, M-1$.

C. Asymptotic Characterization of BICM

The BICM-GMI is an achievable rate for BICM [18] and is one of the key information-theoretic quantities used to analyze BICM systems.⁴ For any P_X and \mathbf{l} , the BICM-GMI is defined as [24, eq. (24)]

$$\tilde{I}_{P_X, \mathbf{l}}(\rho) \triangleq m I_{P_X}(\rho) - \sum_{k=1}^m \sum_{b \in \mathcal{B}} P_{Q_k}(b) I_{P_{X_{k,b}}}(\rho) \quad (50)$$

⁴Note that even though the BICM-GMI is fully determined by the bit probabilities $[P_{Q_1}(0), P_{Q_2}(0), \dots, P_{Q_m}(0)]$, we express it as a function of the input distribution P_X in (38).

and twice its derivative as⁵

$$\tilde{M}_{P_X, \mathbf{l}}(\rho) \triangleq 2 \frac{d\tilde{I}_{P_X, \mathbf{l}}(\rho)}{d\rho} \quad (51)$$

$$= m M_{P_X}(\rho) - \sum_{k=1}^m \sum_{b \in \mathcal{B}} P_{Q_k}(b) M_{P_{X_{k,b}}}(\rho). \quad (52)$$

For the particular case $P_X = P_X^u$ we obtain

$$\tilde{I}_{P_X^u, \mathbf{l}}(\rho) \triangleq m I_{P_X^u}(\rho) - \frac{1}{2} \sum_{k=1}^m \sum_{b \in \mathcal{B}} I_{P_{X_{k,b}}^u}(\rho), \quad (53)$$

$$\tilde{M}_{P_X^u, \mathbf{l}}(\rho) \triangleq m M_{P_X^u}(\rho) - \frac{1}{2} \sum_{k=1}^m \sum_{b \in \mathcal{B}} M_{P_{X_{k,b}}^u}(\rho). \quad (54)$$

Like the MI, the BICM-GMI also tends to H_{P_X} as ρ tends to infinity. The following theorem shows how fast $\tilde{I}_{P_X, \mathbf{l}}(\rho)$ converges to H_{P_X} .

Theorem 4: For any P_X and \mathbf{l}

$$\lim_{\rho \rightarrow \infty} \frac{H_{P_X} - \tilde{I}_{P_X, \mathbf{l}}(\rho)}{Q(\sqrt{\rho}d/2)} = \pi \sum_{k=1}^m \left(N_{P_X} - \sum_{b \in \mathcal{B}} P_{Q_k}(b) N_{P_{X_{k,b}}} \right). \quad (55)$$

where N_{P_X} and $N_{P_{X_{k,b}}}$ are given in (19) and (42), respectively.

Proof: The proof is given in Appendix C. ■

The following theorem gives an asymptotic expression for $\tilde{M}_{P_X, \mathbf{l}}(\rho)$.

Theorem 5: For any P_X and \mathbf{l}

$$\lim_{\rho \rightarrow \infty} \frac{\tilde{M}_{P_X, \mathbf{l}}(\rho)}{Q(\sqrt{\rho}d/2)} = \frac{d^2}{4} \pi \sum_{k=1}^m \left(N_{P_X} - \sum_{b \in \mathcal{B}} P_{Q_k}(b) N_{P_{X_{k,b}}} \right). \quad (56)$$

where N_{P_X} and $N_{P_{X_{k,b}}}$ are given in (19) and (42), respectively.

Proof: By using (52) and Theorem 2. ■

In analogy to (8)–(9), we define the BEP⁶ as

$$B_{P_X, \mathbf{l}}(\rho) \triangleq \frac{1}{m} \sum_{k=1}^m \Pr\{\hat{Q}_k^{\text{MAP}}(Y) \neq Q_k\} \quad (57)$$

where Q_k is the transmitted bit and

$$\hat{Q}_k^{\text{MAP}}(y) \triangleq \underset{b \in \mathcal{B}}{\operatorname{argmax}} P_{Q_k|Y}(b|y) \quad (58)$$

is the decision made by a MAP bit demapper. The next theorem gives an asymptotic expression for the BEP in (57)–(58).

Theorem 6: For any P_X and \mathbf{l}

$$\lim_{\rho \rightarrow \infty} \frac{B_{P_X, \mathbf{l}}(\rho)}{Q(\sqrt{\rho}d/2)} = E_{P_X, \mathbf{l}} \quad (59)$$

where $E_{P_X, \mathbf{l}}$ is given in (45).

Proof: The proof is given in Appendix D. ■

⁵Since the BICM-GMI is not an MI, its derivative is not an MMSE [31]. We thus avoid using the name MMSE, although we do use an MMSE-like notation $\tilde{M}_{P_X, \mathbf{l}}(\rho)$.

⁶Note that (57) is the BEP averaged over the m bit positions, in contrast to the BICM-GMI in (50), that is a sum of m bit-wise MIs.

Similarly to (26)–(28), we can use Theorems 4–6 to show that, at high SNR, the BICM-GMI, twice its derivative, and the BEP behave as in (60)–(62) (see next page).

The results in (60)–(62) show that in the high-SNR regime, the BICM-GMI, its derivative, and the BEP have the same asymptotic behavior⁷, i.e., they are all proportional to a Gaussian Q-function that depends only on the MED of the constellation.

For a uniform input distribution, Theorems 4–6 particularize to the following result.

Corollary 2: For any \mathcal{X} and \mathbf{l} and a uniform input distribution

$$\lim_{\rho \rightarrow \infty} \frac{\log M - \tilde{I}_{P_X^u, \mathbf{l}}(\rho)}{Q(\sqrt{\rho}d/2)} = \pi \frac{C_{\mathcal{X}, \mathbf{l}}}{M}, \quad (63)$$

$$\lim_{\rho \rightarrow \infty} \frac{\tilde{M}_{P_X^u, \mathbf{l}}(\rho)}{Q(\sqrt{\rho}d/2)} = \frac{d^2}{4} \pi \frac{C_{\mathcal{X}, \mathbf{l}}}{M}, \quad (64)$$

$$\lim_{\rho \rightarrow \infty} \frac{B_{P_X^u, \mathbf{l}}(\rho)}{Q(\sqrt{\rho}d/2)} = \frac{C_{\mathcal{X}, \mathbf{l}}}{mM} \quad (65)$$

where $C_{\mathcal{X}, \mathbf{l}}$ is given in (44).

Proof: The expression in (63) follows from Theorem 4 combined with (21), (43), (47), and $P_{Q_k}(b) = 1/2$, the one in (64) from Theorem 5 combined with (21), (43), (47), and $P_{Q_k}(b) = 1/2$, and the one in (65) from Theorem 6 and (46). ■

The results in Corollary 2 indicate that, for a uniform input distribution, a maximization of the BICM-GMI is equivalent to a minimization of both its derivative and the BEP. The asymptotic results for BICM are summarized in Table II.

D. Optimality of Gray Codes

To study the asymptotic behavior of the BICM-GMI for different labelings \mathbf{l} , we introduce the two functions

$$\mathcal{K}_{P_X, \mathbf{l}}^I(\rho) \triangleq \frac{H_{P_X} - \tilde{I}_{P_X, \mathbf{l}}(\rho)}{H_{P_X} - I_{P_X}(\rho)} \quad (66)$$

$$\mathcal{K}_{P_X, \mathbf{l}}^M(\rho) \triangleq \frac{\tilde{M}_{P_X, \mathbf{l}}(\rho)}{M_{P_X}(\rho)}. \quad (67)$$

Noting that $\tilde{I}_{P_X, \mathbf{l}}(\rho) \leq I_{P_X}(\rho)$ [16, eq. (16)], [23, Theorem 5], we have

$$\mathcal{K}_{P_X, \mathbf{l}}^I(\rho) \geq 1. \quad (68)$$

We further define

$$R_{P_X, \mathbf{l}} \triangleq \lim_{\rho \rightarrow \infty} \mathcal{K}_{P_X, \mathbf{l}}^I(\rho) \quad (69)$$

$$= \lim_{\rho \rightarrow \infty} \mathcal{K}_{P_X, \mathbf{l}}^M(\rho) \quad (70)$$

where (70) follows from L'Hôpital's rule. Theorems 1 and 4 yield

$$R_{P_X, \mathbf{l}} = \frac{\sum_{k=1}^m (N_{P_X} - \sum_{b \in \mathcal{B}} P_{Q_k}(b) N_{P_{X_{k,b}}})}{N_{P_X}} \quad (71)$$

and because of (68)

$$R_{P_X, \mathbf{l}} \geq 1. \quad (72)$$

⁷Disregarding the “offset” H_{P_X} in (60).

TABLE II
SUMMARY OF ASYMPTOTICS OF THE BICM-GMI, TWICE ITS DERIVATIVE, AND THE BEP.

Input Distribution	P_X	P_X^u
$\lim_{\rho \rightarrow \infty} \frac{H_{P_X} - \tilde{I}_{P_X, \mathbf{l}}(\rho)}{Q(\sqrt{\rho}d/2)}$	$\pi \sum_{k=1}^m \left(N_{P_X} - \sum_{b \in \mathcal{B}} P_{Q_k}(b) N_{P_{X_{k,b}}} \right)$	$\pi \frac{C_{\mathcal{X}, \mathbf{l}}}{M}$
$\lim_{\rho \rightarrow \infty} \frac{\tilde{M}_{P_X, \mathbf{l}}(\rho)}{Q(\sqrt{\rho}d/2)}$	$\frac{d^2}{4} \pi \sum_{k=1}^m \left(N_{P_X} - \sum_{b \in \mathcal{B}} P_{Q_k}(b) N_{P_{X_{k,b}}} \right)$	$\frac{d^2}{4} \pi \frac{C_{\mathcal{X}, \mathbf{l}}}{M}$
$\lim_{\rho \rightarrow \infty} \frac{B_{P_X, \mathbf{l}}(\rho)}{Q(\sqrt{\rho}d/2)}$	$E_{P_X, \mathbf{l}}$	$\frac{C_{\mathcal{X}, \mathbf{l}}}{mM}$

In the rest of this section, we study $R_{P_X, \mathbf{l}}$ in (71) for a uniform input distribution P_X^u . With a slight abuse of notation, we will refer to $R_{P_X^u, \mathbf{l}}$ as $R_{\mathcal{X}, \mathbf{l}}$. We shall say that, for a constellation \mathcal{X} and a uniform input distribution, a labeling \mathbf{l} is asymptotically optimal (AO) in terms of BICM-GMI if it satisfies $R_{\mathcal{X}, \mathbf{l}} = 1$. Intuitively, this can be understood as follows. The condition $R_{\mathcal{X}, \mathbf{l}} = 1$ guarantees that $\lim_{\rho \rightarrow \infty} K_{P_X, \mathbf{l}}^1(\rho) = 1$, and because of (66), an AO labeling is then a binary labeling for which the BICM-GMI approaches H_{P_X} as fast as the MI does for the same constellation \mathcal{X} .

Corollary 3: For any labeling \mathbf{l} and constellation \mathcal{X} , $R_{\mathcal{X}, \mathbf{l}}$ can be expressed as

$$R_{\mathcal{X}, \mathbf{l}} = \frac{C_{\mathcal{X}, \mathbf{l}}}{A_{\mathcal{X}}}. \quad (73)$$

Proof: Follows by combining (66) and (69) with (29) and (64). ■

Example 4: For an MPAM constellation, $R_{\mathcal{E}, \mathbf{l}}$ can be expressed using (18) and (73) as

$$R_{\mathcal{E}, \mathbf{l}} = \frac{C_{\mathcal{E}, \mathbf{l}}}{2(M-1)}. \quad (74)$$

If the labeling is the NBC, we further use (49) to obtain

$$R_{\mathcal{E}, \mathbf{l}_{\text{NBC}}} = \frac{2M - m - 2}{M - 1} \quad (75)$$

which fulfills $R_{\mathcal{E}, \mathbf{l}_{\text{NBC}}} > 1$ for $m > 1$. We thus conclude that the NBC for MPAM is not an AO labeling.

The following theorem demonstrates that GCs are AO at high SNR, and thus, it proves the conjecture of the optimality of GCs in terms of BICM-GMI [16, Sec. III-C].

Theorem 7: For any constellation \mathcal{X} and a uniform input distribution, a labeling is AO if and only if it is a GC.

Proof: For any GC, all the pairs of constellation points at MED are at Hamming distance one. Thus, (48) holds with equality, and therefore, $R_{\mathcal{X}, \mathbf{l}} = 1$. This completes the “if”

part of the proof. The “only if” part follows from that for any non-GC, there is at least one pair of constellation points at Hamming distance larger than one, thus, $C_{\mathcal{X}, \mathbf{l}} > A_{\mathcal{X}}$, and therefore, $R_{\mathcal{X}, \mathbf{l}} > 1$. ■

Remark 3: The results about the optimality of GCs directly extends to multidimensional constellations that are constructed as direct products of one-dimensional constellations, provided that the labeling is generated via an ordered direct product of GCs. This construction of constellation and labelings was formally used in e.g., [23, Theorem 15].

Remark 4: The NBC was shown in Example 4 not to be AO for an MPAM constellation. However, if the NBC is used with a nonequally spaced constellation, it may become AO. This is the case for example if the NBC is used with the constellation in Example 1, in which case the NBC is a GC.

We next derive an upper bound on $R_{\mathcal{X}, \mathbf{l}}$.

Theorem 8: For any one-dimensional constellation and any labeling \mathbf{l}

$$C_{\mathcal{X}, \mathbf{l}} \leq \min(mA_{\mathcal{X}}, (m-1)A_{\mathcal{X}} + M) \quad (76)$$

and thus,

$$R_{\mathcal{X}, \mathbf{l}} \leq \frac{\min(mA_{\mathcal{X}}, (m-1)A_{\mathcal{X}} + M)}{A_{\mathcal{X}}}. \quad (77)$$

Proof: We note that for any labeling there are exactly $M/2$ pairs of labels at Hamming distance m . Because of this, at most $M/2$ pairs of constellation points at MED can each differ in exactly m bits, which can be the case only if $A_{\mathcal{X}} \leq M$. This case gives $C_{\mathcal{X}, \mathbf{l}} \leq mA_{\mathcal{X}}$. If there are more than $M/2$ pairs of constellation points at MED, i.e., $A_{\mathcal{X}} > M$, $M/2$ pairs can differ in m bits and the remaining $(A_{\mathcal{X}} - M)/2$ pairs can differ in at most $m-1$ bits, which gives $C_{\mathcal{X}, \mathbf{l}} \leq mM + (m-1)(A_{\mathcal{X}} - M) = (m-1)A_{\mathcal{X}} + M$. The expression in (77) follows from (76) and (73). ■

$$\tilde{I}_{P_X, \mathbf{l}}(\rho) \approx H_{P_X} - \pi \sum_{k=1}^m \left(N_{P_X} - \sum_{b \in \mathcal{B}} P_{Q_k}(b) N_{P_{X_{k,b}}} \right) Q\left(\frac{\sqrt{\rho}d}{2}\right), \quad (60)$$

$$\tilde{M}_{P_X, \mathbf{l}}(\rho) \approx \frac{d^2}{4} \pi \sum_{k=1}^m \left(N_{P_X} - \sum_{b \in \mathcal{B}} P_{Q_k}(b) N_{P_{X_{k,b}}} \right) Q\left(\frac{\sqrt{\rho}d}{2}\right), \quad (61)$$

$$B_{P_X, \mathbf{l}}(\rho) \approx E_{P_X, \mathbf{l}} Q\left(\frac{\sqrt{\rho}d}{2}\right). \quad (62)$$

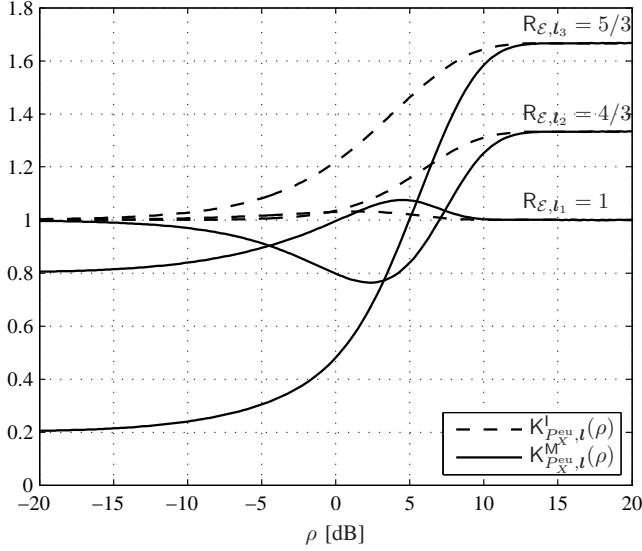


Fig. 5. Functions $K_{P_X^{\text{eu}},l}^l(\rho)$ (dashed lines) and $K_{P_X^{\text{eu}},l}^M(\rho)$ (solid lines) for 4PAM (normalized to $E_s = 1$) and different labelings. The values of $R_{\mathcal{E},l}$ in (74) are also shown.

For an $MPAM$ constellation, using (18), Theorem 8 yields

$$R_{\mathcal{E},l} \leq m - \frac{M-2}{2M-2}. \quad (78)$$

Example 5: In Fig. 5, we show the functions $K_{P_X^{\text{eu}},l}^l(\rho)$ and $K_{P_X^{\text{eu}},l}^M(\rho)$ in (66) and (67), respectively, for a 4PAM constellation with a uniform input distribution ($P_X = P_X^{\text{eu}}$, $A_{\mathcal{X}} = 6$) and the three labelings that give different BICM-GMI: $l_1 = [0, 1, 3, 2]$, $l_2 = [0, 1, 2, 3]$ and $l_3 = [0, 3, 2, 1]$. The values of $R_{\mathcal{E},l}$ in (74) are also shown. Unlike the BICM-GMI curves usually plotted in previous works (see [19, Fig. 3] and [31, Fig. 1]), the functions $K_{P_X^{\text{eu}},l}^l(\rho)$ and $K_{P_X^{\text{eu}},l}^M(\rho)$ allow us to study different labelings at high SNR. Moreover, $K_{P_X^{\text{eu}},l}^M(\rho)$ also allows us to study different labelings at low SNR: Fig. 5 shows that the NBC (i.e., l_2), is the binary labeling for 4PAM that gives the largest value for $\tilde{M}_{P_X^{\text{eu}},l}(\rho)$ as ρ tends to zero, cf. [22], [23, Theorem 14].⁸ Observe that the GC (i.e., l_1), as expected, gives $R_{\mathcal{E},l_1} = 1$, and that l_3 achieves the upper bound in (78), i.e., $R_{\mathcal{E},l_3} = 5/3$.

Example 6: In Fig. 6 we show the function $K_{P_X^{\text{eu}},l}^M(\rho)$ in (67) for 8PAM ($P_X = P_X^{\text{eu}}$, $A_{\mathcal{X}} = 14$) and all the 458 labelings that give a different BICM-GMI [25]. In this figure, 12 possible values of $R_{\mathcal{E},l}$ in (74) are clearly visible, which coincide with the results in [25, Fig. 3].⁹ Using (65), the 12 values of $R_{\mathcal{E},l}$ in Fig. 6 also translate into 12 different asymptotic BEP curves, which were recently reported in [32, Fig. 4]. The value $R_{\mathcal{E},l_{\text{NBC}}}$ obtained using (75) is also shown. A careful examination of Fig. 6 reveals that there are three labelings minimizing $R_{\mathcal{E},l}$. These are the three nonequivalent GCs (in terms of BEP) [29, Table I]: the BRGC $l = [0, 1, 3, 2, 6, 7, 5, 4]$,

⁸The relationship between the coefficient α determining the low-SNR regime for a zero-mean constellation with a uniform input distribution [23, eq. (47)] is $\alpha \log 2 = \lim_{\rho \rightarrow 0} K_{P_X^{\text{eu}},l}^M(\rho)$ (see also [13, eq. (86)]).

⁹Further note that $\lim_{\rho \rightarrow 0} K_{P_X^{\text{eu}},l}^M(\rho)$ reveals the 72 classes of labelings reported in [23, Fig. 6 (a)].

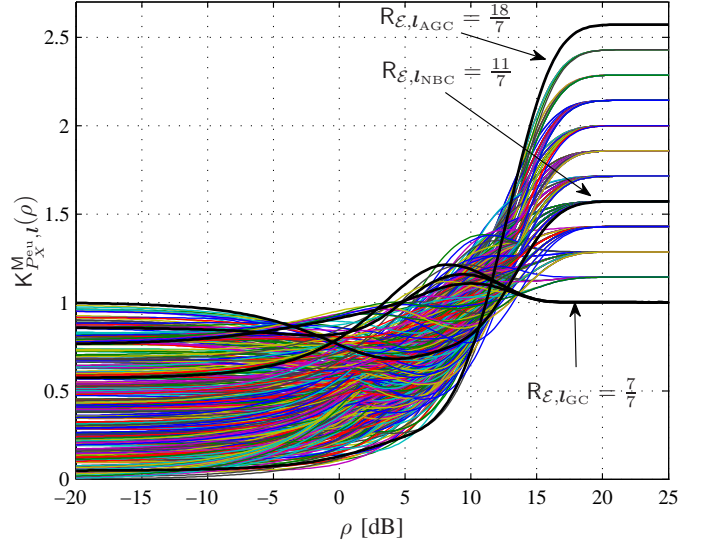


Fig. 6. Function $K_{P_X^{\text{eu}},l}^M(\rho)$ for the 458 labelings that give a different BICM-GMI for 8PAM (normalized to $E_s = 1$). The values of $R_{\mathcal{E},l}$ in (74) for the three nonequivalent GCs, the NBC, and the AGC are also shown.

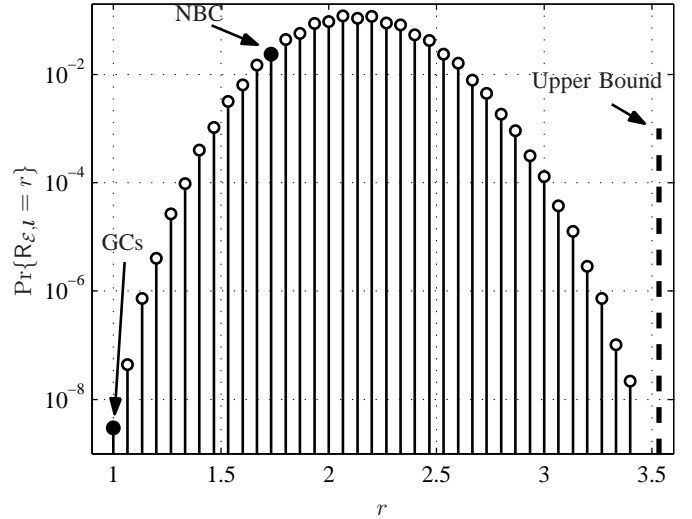


Fig. 7. Approximated $\Pr\{R_{\mathcal{E},l} = r\}$ using 10^9 randomly generated labelings for 16PAM (normalized to $E_s = 1$). For GCs $R_{\mathcal{E},l_{\text{GC}}} = 1$ and for the NBC $R_{\mathcal{E},l_{\text{NBC}}} = 26/15$. The upper bound in (78) is also shown.

$l = [0, 1, 3, 2, 6, 4, 5, 7]$, and $l = [0, 1, 3, 7, 5, 4, 6, 2]$.

Example 7: Motivated by [23, Fig. 6], we present in Fig. 7 an approximation for the PMF $\Pr\{R_{\mathcal{X},l} = r\}$ for 16PAM obtained by randomly generating 10^9 labelings. This figure shows that most of the possible labelings are not Gray. For $M = 16$, we obtain $R_{\mathcal{E},l_{\text{NBC}}} = 26/15$, cf. (75), which is highlighted in Fig. 7. The upper bound in (78) is also shown. In the next section, we will show how to construct a labeling that achieves this upper bound.

E. Anti-Gray Codes

In the previous section, we have seen that GCs minimize $R_{\mathcal{X},l}$. In this section, we show that, for $MPAM$ constellations,

it is always possible to construct a labeling that maximizes $R_{\mathcal{E},l}$, i.e., a labeling that achieves the upper bound in (78).

We define the set of all possible values that $C_{\mathcal{X},l}$ can take as $\mathcal{C}_{\mathcal{X}}$, where

$$|\mathcal{C}_{\mathcal{X}}| \leq \frac{1}{2} \min \{(m-1)A_{\mathcal{X}} + 2, (m-2)A_{\mathcal{X}} + M + 2\} \quad (79)$$

which follows from the fact that $C_{\mathcal{X},l}$ is an even integer bounded by (48) and (76).

The expression in (79) is an upper bound on the number of classes of labelings with different high-SNR behavior in terms of BICM-GMI or BEP. For the particular case of $\mathcal{X} = \mathcal{E}$, by using (18) in (79), we obtain

$$|\mathcal{C}_{\mathcal{E}}| \leq mM - \frac{3M}{2} - m + 3. \quad (80)$$

For 4PAM $|\mathcal{C}_{\mathcal{E}}| \leq 3$ and for 8PAM $|\mathcal{C}_{\mathcal{E}}| \leq 12$, which coincides with the 3 and 12 classes for high SNR shown in Fig. 5 and Fig. 6, respectively. For 16PAM, the upper bound (80) indicates that $|\mathcal{C}_{\mathcal{E}}| \leq 39$, however, the results in Fig. 7 show only 37 classes. This raises the question of the tightness of the bound in (80) (or equivalently, the upper bound in (78)), which we address in what follows.

The anti-Gray code (AGC) of order $m \geq 2$ is defined by the $M \times m$ binary matrix \mathbf{W}_m , where the i th row is the binary label for x_i , where $\mathbf{W}_1 = [0, 1]^T$, and where the following steps construct \mathbf{W}_m from \mathbf{W}_{m-1} :

- Step 1 Reverse the $M/2$ rows in \mathbf{W}_{m-1} , and append them under \mathbf{W}_{m-1} to construct a new matrix \mathbf{W}'_m with M rows and $m-1$ columns.
- Step 2 Append the length M column vector $[0, 1, 0, 1, \dots, 0, 1]^T$ to the left of \mathbf{W}'_m to create \mathbf{W}''_m , with M rows and m columns.
- Step 3 Negate all bits in the lower half of \mathbf{W}''_m to obtain \mathbf{W}_m .

The recursive construction described above is illustrated in Fig. 8 for $m = 2$ and $m = 3$. The following lemma shows that this construction indeed leads to a valid labeling.

Lemma 4: All the rows in \mathbf{W}_m are unique, and thus, the AGC is a valid labeling.

Proof: Consider the above construction of an AGC. \mathbf{W}_1 is a valid labeling in the sense that all rows are unique. Because of Step 1 every odd row in the upper half of \mathbf{W}'_m is identical to an even row in the lower half of \mathbf{W}'_m , which directly implies that all rows of \mathbf{W}''_m in Step 2 are unique. Thus, \mathbf{W}''_m is a valid labeling. Moreover, every odd row in the lower half of \mathbf{W}''_m differs in m bits compared to the row below. Inverting all the bits in these $M/2$ rows is therefore equivalent to swapping every odd row in the lower half of \mathbf{W}''_m with the row below, which makes \mathbf{W}_m a valid labeling with M unique rows. ■

The next theorem proves that, at high SNR, the AGC is the best binary labeling for MPAM constellations.

Theorem 9: For $\mathcal{X} = \mathcal{E}$, the AGC achieves the upper bound in (78), i.e.,

$$R_{\mathcal{E},l_{\text{AGC}}} = m - \frac{M-2}{2M-2}. \quad (81)$$

Proof: Let $H_m = C_{\mathcal{E},l_{\text{AGC}}}$ denote twice the sum of the Hamming distances between all adjacent rows in \mathbf{W}_m , and

let H'_m and H''_m denote the same quantity for \mathbf{W}'_m and \mathbf{W}''_m , respectively. Steps 1 and 2 give $H'_m = 2H_{m-1}$ and $H''_m = H'_m + 2(M-1)$. It then follows that $H_m = H''_m - 2 + 2(m-1)$, since row $M/2$ and row $M/2+1$ in \mathbf{W}''_m differ in only one bit and therefore the same rows in \mathbf{W}_m differ in $m-1$ bits. This gives $H_m = 2H_{m-1} + 2(M+m-3)$, which combined with $H_1 = 2$ can be shown to give $H_m = 2(mM - m - \frac{M}{2} + 1)$. Together with (74), this completes the proof. ■

The labeling l_3 in Example 5 and Fig. 5 (i.e., \mathbf{W}_2 in Fig. 8) is the AGC for 4PAM with $R_{\mathcal{E},l} = 5/3$ given by (81). For 8PAM, the AGC is $l_{\text{AGC}} = [0, 7, 2, 5, 6, 1, 4, 3]$ (\mathbf{W}_3 in Fig. 8), whose corresponding function $K_{P_{\mathcal{X}}^{\text{av}}, l_{\text{AGC}}}(\rho)$ is shown in Fig. 6, with $R_{\mathcal{E},l} = 18/7$.

For $M = 16$, the labeling that maximizes $R_{\mathcal{E},l}$ ($R_{\mathcal{E},l} = 106/30 \approx 3.53$) is the AGC \mathbf{W}_4 (as shown by Theorem 9), which can be constructed as described before. It can be further shown that the labeling with the second largest $R_{\mathcal{E},l}$ ($R_{\mathcal{E},l} = 104/30 \approx 3.47$) can be constructed by reversing the order of the three first rows of the AGC \mathbf{W}_4 . This demonstrates that for 16PAM all 39 classes are indeed possible. The last two classes are not shown in Fig. 7 because the total number of labelings in this case is $16! \approx 2.1 \cdot 10^{13}$ (without discarding trivial operations), so randomly generating 10^9 labelings only covers a small fraction of all possible labelings.

V. CONCLUSIONS

In this paper, we studied discrete constellations with arbitrary input distributions over the scalar AWGN channel in the high-SNR regime and derived exact asymptotic expressions for key quantities in information theory, estimation theory, and communication theory. The six quantities we studied are the MI, MMSE, SEP, the BICM-GMI, its derivative, and BEP. Our results show that, as SNR tends to infinity and disregarding offsets, all these quantities are proportional to $Q(\sqrt{\rho d}/2)$, where d is the MED of the constellation. These results show asymptotic equivalences between all these quantities as well as the importance of the Gaussian Q-function.

For a uniform input distribution, the proportionality constants for the MI, SEP, and MMSE were found to be a function of the MED of the constellation and the number of pairs of constellation points at MED only, and thus, the constellation that maximizes the MI in the high-SNR regime is the same that minimizes both the SEP and the MMSE.

We then applied our results to the problem of binary labelings for BICM. By characterizing the high-SNR behavior of the BICM-GMI, asymptotically optimal binary labelings were found, and the long-standing conjecture that Gray codes are optimal at high SNR was proved. We also proved that there always exists an anti-Gray code for MPAM constellations, which is the labeling that has the lowest BICM-GMI at high SNR.

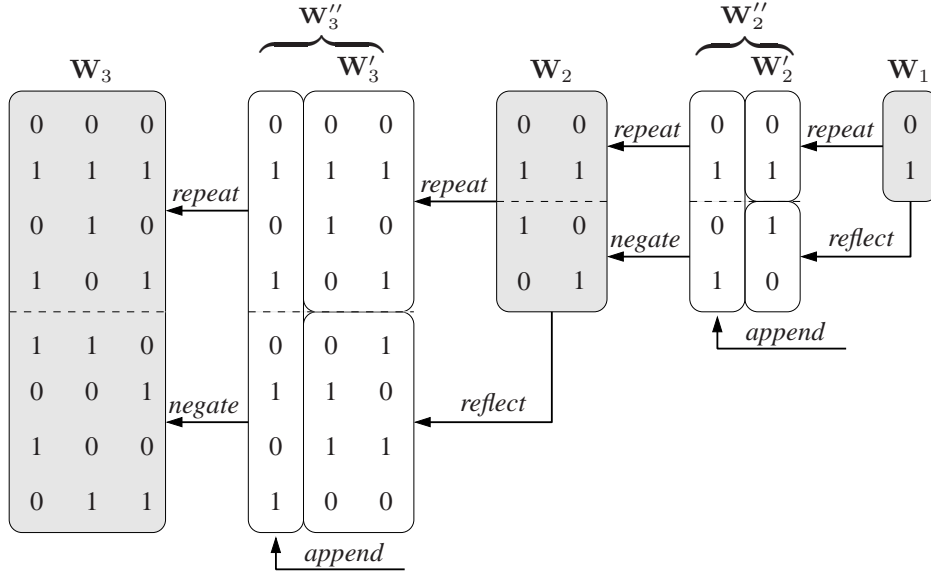


Fig. 8. Proposed recursive construction of an AGC for $m = 2$ and $m = 3$.

APPENDIX A PROOF OF THEOREM 1

We start by upper and lower bounding the Q-function via [33, Prop. 19.4.2]

$$\left(1 - \frac{1}{x^2}\right) G(x) \leq Q(x) \leq G(x), \quad x > 0 \quad (82)$$

where

$$G(x) \triangleq \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (83)$$

It follows that

$$\lim_{x \rightarrow \infty} \frac{G(x)}{Q(x)} = 1. \quad (84)$$

The numerator of the l.h.s. of (23) can be expressed as

$$H_{P_X} - I_{P_X}(\rho) = \sum_{i \in \mathcal{I}_X} p_i V_i(\rho) \quad (85)$$

where

$$V_i(\rho) \triangleq \int_{-\infty}^{\infty} \frac{e^{-t^2}}{\sqrt{\pi}} \cdot \log \left(\sum_{\delta \in \mathcal{D}_X^{(i)}} \left(B_{P_X}^{(i)}(\delta) \right)^2 e^{-\sqrt{2\rho}t\delta - \frac{\rho\delta^2}{2}} \right) dt. \quad (86)$$

The expressions in (85) and (86) are obtained from [25, eq. (5)] and the definition of entropy.

Combining (84) and (85) yields

$$\lim_{\rho \rightarrow \infty} \frac{H_{P_X} - I_{P_X}(\rho)}{Q(\sqrt{\rho}d/2)} = \lim_{\rho \rightarrow \infty} \frac{H_{P_X} - I_{P_X}(\rho)}{G(\sqrt{\rho}d/2)} \quad (87)$$

$$= \sum_{i \in \mathcal{I}_X} p_i \lim_{\rho \rightarrow \infty} \frac{V_i(\rho)}{G(\sqrt{\rho}d/2)}. \quad (88)$$

As will become apparent later, the limit on the right-hand side (r.h.s.) of (88) exists and, hence, so does the limit on the l.h.s.

of (87).

In what follows, we calculate the limit on the r.h.s. of (88). Using (83) and (86), and substituting $r = d\sqrt{\rho}/8$, we obtain

$$\lim_{\rho \rightarrow \infty} \frac{V_i(\rho)}{G(\sqrt{\rho}d/2)} = 2 \left(\lim_{r \rightarrow \infty} F_i^-(r) + \lim_{r \rightarrow \infty} F_i^+(r) \right) \quad (89)$$

where

$$F_i^-(r) \triangleq \int_{-\infty}^0 r e^{r^2 - t^2} \cdot \log \left(\sum_{\delta \in \mathcal{D}_X^{(i)}} \left(B_{P_X}^{(i)}(\delta) \right)^2 e^{-4rt\frac{\delta}{d} - 4r^2\frac{\delta^2}{d^2}} \right) dt \quad (90)$$

and

$$F_i^+(r) \triangleq \int_0^{\infty} r e^{r^2 - t^2} \cdot \log \left(\sum_{\delta \in \mathcal{D}_X^{(i)}} \left(B_{P_X}^{(i)}(\delta) \right)^2 e^{-4rt\frac{\delta}{d} - 4r^2\frac{\delta^2}{d^2}} \right) dt. \quad (91)$$

We begin with the first limit on the r.h.s. of (89). Using the substitution $t = u/r - r$, we express $F_i^-(r)$ in (90) as

$$F_i^-(r) = \int_{-\infty}^{r^2} e^{2u - \frac{u^2}{r^2}} \cdot \log \left(\sum_{\delta \in \mathcal{D}_X^{(i)}} \left(B_{P_X}^{(i)}(\delta) \right)^2 e^{-4u\frac{\delta}{d} - 4r^2U(\delta)} \right) du \quad (92)$$

where

$$U(\delta) \triangleq \frac{\delta}{d} \left(\frac{\delta}{d} - 1 \right) \quad (93)$$

and $U(\delta) \geq 0, \forall \delta \in \mathcal{D}_X$. Defining

$$f_i^-(r, u) \triangleq h(r^2 - u) \cdot e^{2u - \frac{u^2}{r^2}} \cdot \log \left(1 + \sum_{\delta \in \mathcal{D}_i^*} \left(B_{P_X}^{(i)}(\delta) \right)^2 e^{-4u \frac{\delta}{d} - 4r^2 U(\delta)} \right) \quad (94)$$

with $\mathcal{D}_i^* \triangleq \mathcal{D}_X^{(i)} \setminus \{0\}$ and $h(x)$ being the Heaviside's step function (i.e., $h(x) = 1$ if $x \geq 0$ and $h(x) = 0$ if $x < 0$), $F_i^-(r)$ in (92) can be written as

$$F_i^-(r) = \int_{-\infty}^{\infty} f_i^-(r, u) du. \quad (95)$$

Note that $f_i^-(r, u)$ is a nonnegative function, and because $U(d) = 0$,

$$\lim_{r \rightarrow \infty} f_i^-(r, u) = e^{2u} \log \left(1 + \left(B_{P_X}^{(i)}(d) \right)^2 e^{-4u} \right), \quad u \in \mathbb{R}. \quad (96)$$

We will further show that, for every $r > 0$, $f_i^-(r, u)$ is uniformly bounded by some integrable function $g_i^-(u)$ that is independent of r (see Lemma 5 ahead). To compute the first limit on the r.h.s. of (89), we can thus use Lebesgue's Dominated Convergence Theorem [34, Theorem 1.34] to obtain

$$\lim_{r \rightarrow \infty} F_i^-(r) = \lim_{r \rightarrow \infty} \int_{-\infty}^{\infty} f_i^-(r, u) du \quad (97)$$

$$= \int_{-\infty}^{\infty} \lim_{r \rightarrow \infty} f_i^-(r, u) du \quad (98)$$

$$= \int_{-\infty}^{\infty} e^{2u} \log \left(1 + \left(B_{P_X}^{(i)}(d) \right)^2 e^{-4u} \right) du \quad (99)$$

$$= \frac{\pi}{2} B_{P_X}^{(i)}(d) \quad (100)$$

where (99) is obtained from (96), and (100) follows from using the substitution $x = e^{2u}$ together with [35, eq. (4.222.1)].

It thus remains to show that $f_i^-(r, u)$ is uniformly bounded by some integrable function $g_i^-(u)$ that is independent of r . We do this in the following lemma.

Lemma 5: For any $r > 0$ we have

$$0 \leq f_i^-(r, u) \leq g_i^-(u), \quad u \in \mathbb{R} \quad (101)$$

where $g_i^-(u)$ is defined as

$$g_i^-(u) \triangleq \begin{cases} e^{2u} \log \left(\frac{M}{p_i} e^{-4u \hat{d}^2 / d^2} \right), & u < 0 \\ e^{2u} \log \left(1 + \frac{M-1}{p_i} e^{-4u} \right), & u \geq 0 \end{cases} \quad (102)$$

and satisfies $g_i^-(u) \geq 0, u \in \mathbb{R}$ and

$$\int_{-\infty}^{\infty} g_i^-(u) du < \infty \quad (103)$$

and where \hat{d} is given by (13).

Proof: We use $e^{-\frac{u^2}{r^2}} \leq 1$ and $\left(B_{P_X}^{(i)}(\delta) \right)^2 < 1/p_i$ to

upper-bound (94) as

$$f_i^-(r, u) \leq h(r^2 - u) e^{2u} \log \left(1 + \sum_{\delta \in \mathcal{D}_i^*} \frac{e^{-4u \frac{\delta}{d} - 4r^2 U(\delta)}}{p_i} \right) \quad (104)$$

$$\leq e^{2u} \log \left(1 + \sum_{\delta \in \mathcal{D}_i^*} \frac{e^{-4u \frac{\delta^2}{d^2}}}{p_i} \right) \quad (105)$$

where to pass from (104) to (105) we used $e^{-4r^2 U(\delta)} \leq e^{-4u U(\delta)}$ for $u \leq r^2$ (because $U(\delta) \geq 0$) and that the r.h.s. of (105) is nonnegative for $u < r^2$.

For $u \geq 0$, we have

$$f_i^-(r, u) \leq e^{2u} \log \left(1 + \frac{M-1}{p_i} e^{-4u} \right) \quad (106)$$

which is obtained by replacing δ^2 in (105) by d^2 . For $u < 0$, (105) is upper-bounded by

$$f_i^-(r, u) \leq e^{2u} \log \left(\frac{M}{p_i} e^{-4u \hat{d}^2 / d^2} \right) \quad (107)$$

which is obtained by using $1 < 1/p_i < e^{-4u \hat{d}^2 / d^2} / p_i$ and by replacing δ^2 in (105) by \hat{d}^2 . The proof of (103) follows from

$$\int_{-\infty}^{\infty} g_i^-(u) du = \int_{-\infty}^0 g_i^-(u) du + \int_0^{\infty} g_i^-(u) du \quad (108)$$

where

$$\int_{-\infty}^0 g_i^-(u) du = \frac{1}{2} \log \frac{M}{p_i} + \frac{\hat{d}^2}{d^2} \quad (109)$$

and

$$\begin{aligned} \int_0^{\infty} g_i^-(u) du &\leq \int_{-\infty}^{\infty} e^{2u} \log \left(1 + \frac{M-1}{p_i} e^{-4u} \right) du \\ &= \frac{\pi}{2} \sqrt{\frac{M-1}{p_i}} \end{aligned} \quad (110)$$

which follows in analogy to (99)–(100). ■

The second limit on the r.h.s. of (89) can be computed along the same lines by substituting $t = u/r + r$ in (91), which gives

$$\lim_{r \rightarrow \infty} F_i^+(r) = \frac{\pi}{2} B_{P_X}^{(i)}(-d). \quad (111)$$

Combining (100) and (111) with (89) and (88) yields

$$\lim_{\rho \rightarrow \infty} \frac{H_{P_X} - I_{P_X}(\rho)}{Q(\sqrt{\rho}d/2)} = \sum_{i \in \mathcal{I}_X} p_i \pi \left(B_{P_X}^{(i)}(d) + B_{P_X}^{(i)}(-d) \right) \quad (112)$$

which, by (19), is equal to πN_{P_X} . This proves Theorem 1.

APPENDIX B PROOF OF THEOREM 3

Using Bayes' rule, $\hat{X}^{\text{MAP}}(y)$ in (9) is expressed as

$$\hat{X}^{\text{MAP}}(y) = \underset{x \in \mathcal{X}}{\operatorname{argmax}} \{ f_{Y|X}(y|x) P_X(x) \} \quad (113)$$

$$= x_i, \quad \text{if } y \in \mathcal{Y}_i^{\text{MAP}}(\rho), i = 1, \dots, M \quad (114)$$

where $\mathcal{Y}_i^{\text{MAP}}(\rho)$ is the Voronoi region for the symbol x_i defined as

$$\mathcal{Y}_i^{\text{MAP}}(\rho) \triangleq \left\{ y \in \mathbb{R} : \frac{p_i f_{Y|X}(y|x_i)}{p_j f_{Y|X}(y|x_j)} > 1, j \in \mathcal{I}_{\mathcal{X}} \setminus \{i\} \right\} \quad (115)$$

$$= \{y \in \mathbb{R} : \beta_{i-1}^{\text{MAP}}(\rho) \leq y < \beta_i^{\text{MAP}}(\rho)\} \quad (116)$$

and $\beta_i^{\text{MAP}}(\rho)$, for $i = 1, \dots, M-1$, are the thresholds of the Voronoi regions $\mathcal{Y}_i^{\text{MAP}}(\rho)$, $\beta_0^{\text{MAP}}(\rho) \triangleq -\infty$ and $\beta_M^{\text{MAP}}(\rho) \triangleq +\infty$. Using (115), the thresholds¹⁰ $\beta_i^{\text{MAP}}(\rho)$ in (116) are found by solving

$$p_i f_{Y|X}(\beta_i^{\text{MAP}}(\rho)|x_i) = p_{i+1} f_{Y|X}(\beta_i^{\text{MAP}}(\rho)|x_{i+1}) \quad (117)$$

which gives

$$\beta_i^{\text{MAP}}(\rho) = \frac{\log(p_{i+1}/p_i)}{\sqrt{\rho}(x_i - x_{i+1})} + \frac{\sqrt{\rho}(x_i + x_{i+1})}{2}. \quad (118)$$

On the other hand, the maximum likelihood (ML) symbol demapper is defined as

$$\begin{aligned} \hat{X}^{\text{ML}}(y) &\triangleq \underset{x \in \mathcal{X}}{\operatorname{argmax}} \{f_{Y|X}(y|x)\} \\ &= \underset{x \in \mathcal{X}}{\operatorname{argmin}} \{(y - \sqrt{\rho}x)^2\}. \end{aligned} \quad (119)$$

From (119) it is clear that

$$\beta_i^{\text{ML}}(\rho) = \frac{\sqrt{\rho}(x_i + x_{i+1})}{2}. \quad (120)$$

Combining (118) and (120), we obtain

$$\lim_{\rho \rightarrow \infty} \frac{\beta_i^{\text{MAP}}(\rho)}{\beta_i^{\text{ML}}(\rho)} = 1 \quad (121)$$

and thus, to prove Theorem 3, it is enough to study the SEP for the ML symbol demapper.

Using (114) and (116), the SEP in (8) is expressed as

$$S_{P_X}(\rho) = \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \Pr\{Y \notin \mathcal{Y}_i^{\text{ML}}(\rho) | X = x_i\} \quad (122)$$

$$\begin{aligned} &= \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \left(Q(\beta_i^{\text{ML}}(\rho) - \sqrt{\rho}x_i) \right. \\ &\quad \left. + Q(\sqrt{\rho}x_i - \beta_{i-1}^{\text{ML}}(\rho)) \right) \end{aligned} \quad (123)$$

$$\begin{aligned} &= \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \left(Q\left(\frac{\sqrt{\rho}(x_{i+1} - x_i)}{2}\right) \right. \\ &\quad \left. + Q\left(\frac{\sqrt{\rho}(x_i - x_{i-1})}{2}\right) \right) \end{aligned} \quad (124)$$

$$= \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \sum_{\delta \in \hat{\mathcal{D}}_i} A_{\mathcal{X}}^{(i)}(\delta) Q\left(\frac{\sqrt{\rho}\delta^2}{2}\right) \quad (125)$$

where $\hat{\mathcal{D}}_i$ is defined as the set of distances between x_i and its (one or two) closest symbols in \mathcal{X} .

¹⁰Throughout this proof we assume, without loss of generality, that ρ is large enough so that $x_i < \beta_i^{\text{MAP}}(\rho) < x_{i+1}$.

Replacing $S_{P_X}(\rho)$ by (125) in the l.h.s. of (25), we obtain

$$\begin{aligned} &\lim_{\rho \rightarrow \infty} \frac{S_{P_X}(\rho)}{Q(\sqrt{\rho}d/2)} \\ &= \lim_{\rho \rightarrow \infty} \frac{\sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \sum_{\delta \in \hat{\mathcal{D}}_i} A_{\mathcal{X}}^{(i)}(\delta) Q(\sqrt{\rho}\delta^2/2)}{Q(\sqrt{\rho}d/2)} \end{aligned} \quad (126)$$

$$= \lim_{\rho \rightarrow \infty} \frac{Q(\sqrt{\rho}d/2) \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \sum_{w \in \mathcal{W}} A_{\mathcal{X}}^{(i)}(wd)}{Q(\sqrt{\rho}d/2)} \quad (127)$$

where to pass from (126) to (127) we consider only the Q -functions with the smallest argument. Combining (127) with (20) completes the proof.

APPENDIX C PROOF OF THEOREM 4

Using the expression for the BICM-GMI (50), we have

$$\begin{aligned} &\lim_{\rho \rightarrow \infty} \frac{H_{P_X} - \tilde{I}_{P_X, \mathcal{I}}(\rho)}{Q(\sqrt{\rho}d/2)} \\ &= \lim_{\rho \rightarrow \infty} \left(m \frac{H_{P_X} - I_{P_X}(\rho)}{Q(\sqrt{\rho}d/2)} \right. \\ &\quad \left. - \sum_{k=1}^m \sum_{b \in \mathcal{B}} P_{Q_k}(b) \frac{H_{P_{X_{k,b}}} - I_{P_{X_{k,b}}}(\rho)}{Q(\sqrt{\rho}d/2)} \right. \\ &\quad \left. - \frac{(m-1)H_{P_X} - \sum_{k=1}^m \sum_{b \in \mathcal{B}} P_{Q_k}(b) H_{P_{X_{k,b}}}}{Q(\sqrt{\rho}d/2)} \right). \end{aligned} \quad (128)$$

The third term on the r.h.s. of (128) is zero because

$$\begin{aligned} &\sum_{k=1}^m \sum_{b \in \mathcal{B}} P_{Q_k}(b) H_{P_{X_{k,b}}} \\ &= - \sum_{k=1}^m \sum_{b \in \mathcal{B}} \sum_{i \in \mathcal{I}_{\mathcal{X}_{k,b}}} P_{Q_k}(b) P_{X|Q_k}(x_i|b) \log P_{X|Q_k}(x_i|b) \end{aligned} \quad (129)$$

$$= - \sum_{k=1}^m \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \log \frac{p_i}{P_{Q_k}(q_{i,k})} \quad (130)$$

$$= mH_{P_X} + \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \sum_{k=1}^m \log P_{Q_k}(q_{i,k}) \quad (131)$$

$$= mH_{P_X} + \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \log \prod_{k=1}^m P_{Q_k}(q_{i,k}) \quad (132)$$

$$= mH_{P_X} - H_{P_X} \quad (133)$$

where to pass from (129) to (130) we used (39), and to pass from (132) to (133) we used (38).

To compute the first two terms on the r.h.s. of (128), we change the order of summation and limit and apply Theorem 1 to each term. This proves Theorem 4.

APPENDIX D PROOF OF THEOREM 6

The BEP in (57) is expressed as

$$B_{P_X, \mathcal{I}}(\rho) = \frac{1}{m} \sum_{k=1}^m \sum_{i \in \mathcal{I}_{\mathcal{X}}} p_i \Pr\{Y \notin \mathcal{Y}_{i,k}^{\text{MAP}}(\rho) | X = x_i\} \quad (134)$$

where $\mathcal{Y}_{i,k}^{\text{MAP}}(\rho)$ is the “bit-wise Voronoi” region for $q_{i,k}$ given by (58), i.e., $\mathcal{Y}_{i,k}^{\text{MAP}}(\rho) \triangleq \{y \in \mathbb{R} : P_{Q_k|Y}(q_{i,k}|y) > P_{Q_k|Y}(\bar{q}_{i,k}|y)\}$. In analogy with the proof of Theorem 3, the asymptotic BEP is obtained by considering the ML bit demapper instead and only symbols x_j at MED from x_i , i.e.,

$$\begin{aligned} & \lim_{\rho \rightarrow \infty} \frac{B_{P_X, \ell}(\rho)}{Q(\sqrt{\rho}d/2)} \\ &= \lim_{\rho \rightarrow \infty} \frac{\frac{1}{m} \sum_{k=1}^m \sum_{i \in \mathcal{I}_X} p_i \Pr\{Y \notin \mathcal{Y}_{i,k}^{\text{ML}}(\rho) | X = x_i\}}{Q(\sqrt{\rho}d/2)} \end{aligned} \quad (135)$$

$$\begin{aligned} &= \lim_{\rho \rightarrow \infty} \frac{\frac{1}{m} \sum_{k=1}^m \sum_{i \in \mathcal{I}_X} p_i \sum_{\substack{j \in \mathcal{I}_X, q_{j,k} \neq q_{i,k} \\ |x_i - x_j| = d}} Q(\sqrt{\rho}d/2)}{Q(\sqrt{\rho}d/2)} \end{aligned} \quad (136)$$

$$= \frac{1}{m} \sum_{k=1}^m \sum_{b \in \mathcal{B}} \sum_{i \in \mathcal{I}_{X_{k,b}}} p_i \sum_{\substack{j \in \mathcal{I}_{X_{k,\bar{b}}} \\ |x_i - x_j| = d}} 1 \quad (137)$$

where to pass from (136) to (137) we expressed the sum over $i \in \mathcal{I}_X$ as a sum over $b \in \mathcal{B}$ and over $i \in \mathcal{I}_{X_{k,b}}$. The expression in (59) is obtained by recognizing the innermost sum in (137) as $A_{X_{k,b}}^{(i)}(-d) + A_{X_{k,\bar{b}}}^{(i)}(+d)$ and by using (45).

REFERENCES

- [1] C. E. Shannon, “A mathematical theory of communications,” *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.
- [2] S. Verdú, “On channel capacity per unit cost,” *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, June 1990.
- [3] —, “Spectral efficiency in the wideband regime,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.
- [4] V. V. Prelov and S. Verdú, “Second-order asymptotics of mutual information,” *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.
- [5] A. Lozano, A. M. Tulino, and S. Verdú, “Optimum power allocation for parallel Gaussian channels with arbitrary input distributions,” *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, July 2006.
- [6] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú, “MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation,” *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.
- [7] Y. Wu and S. Verdú, “MMSE dimension,” *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4857–4879, Aug. 2011.
- [8] D. Duyck, J. Boutros, and M. Moeneclaey, “Precoding for outage probability minimization on block fading channels,” *submitted to IEEE Trans. Inf. Theory*, Mar. 2012, available at <http://arxiv.org/abs/1103.5348>.
- [9] M. R. D. Rodrigues, “On the constrained capacity of multi-antenna fading coherent channels with discrete inputs,” in *IEEE International Symposium on Information Theory (ISIT)*, Saint Petersburg, Russia, July–Aug. 2011.
- [10] —, “Characterization of the constrained capacity of multiple-antenna fading coherent channels driven by arbitrary inputs,” in *IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, July 2012.
- [11] A. G. C. P. Ramos and M. R. D. Rodrigues, “Coherent fading channels driven by arbitrary inputs: Asymptotic characterization of the constrained capacity and related information- and estimation-theoretic quantities,” Oct. 2012, available at <http://arxiv.org/abs/1210.4505>.
- [12] M. R. D. Rodrigues, “Multiple-antenna fading coherent channels with arbitrary inputs: Characterization and optimization of the reliable information transmission rate,” Oct. 2012, available at <http://arxiv.org/abs/1210.6777>.
- [13] D. Guo, S. Shamai (Shitz), and S. Verdú, “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
- [14] D. Guo, “Gaussian channels: Information, estimation and multiuser detection,” Ph.D. dissertation, Princeton University, Princeton, New Jersey, Nov. 2004.
- [15] E. Zehavi, “8-PSK trellis codes for a Rayleigh channel,” *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.
- [16] G. Caire, G. Taricco, and E. Biglieri, “Bit-interleaved coded modulation,” *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
- [17] A. Guillén i Fàbregas, A. Martinez, and G. Caire, “Bit-interleaved coded modulation,” *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.
- [18] A. Martinez, A. Guillén i Fàbregas, and G. Caire, “Bit-interleaved coded modulation revisited: A mismatched decoding perspective,” *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, June 2009.
- [19] C. Stierstorfer and R. F. H. Fischer, “(Gray) Mappings for bit-interleaved coded modulation,” in *IEEE Vehicular Technology Conference (VTC-Spring)*, Dublin, Ireland, Apr. 2007.
- [20] C. Stierstorfer, “A bit-level-based approach to coded multicarrier transmission,” Ph.D. dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, 2009, available at <http://www.opus.ub.uni-erlangen.de/opus/volltexte/2009/1395/>.
- [21] A. Martinez, A. Guillén i Fàbregas, and G. Caire, “Bit-interleaved coded modulation in the wideband regime,” *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5447–5455, Dec. 2008.
- [22] C. Stierstorfer and R. F. H. Fischer, “Asymptotically optimal mappings for BICM with M -PAM and M^2 -QAM,” *IET Electronics Letters*, vol. 45, no. 3, pp. 173–174, Jan. 2009.
- [23] E. Agrell and A. Alvarado, “Optimal alphabets and binary labelings for BICM at low SNR,” *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6650–6672, Oct. 2011.
- [24] —, “Signal shaping for BICM at low SNR,” *IEEE Trans. Inf. Theory (to appear)*, available at <http://arxiv.org/abs/1202.6404>.
- [25] A. Alvarado, F. Brännström, and E. Agrell, “High SNR bounds for the BICM capacity,” in *IEEE Information Theory Workshop (ITW)*, Paraty, Brazil, Oct. 2011.
- [26] J. B. Anderson and A. Svensson, *Coded Modulation Systems*. Springer, 2003.
- [27] A. Guillén i Fàbregas and A. Martinez, “Bit-interleaved coded modulation with shaping,” in *IEEE Information Theory Workshop (ITW)*, Dublin, Ireland, Aug.–Sep. 2010.
- [28] F. Gray, “Pulse code communications,” U. S. Patent 2 632 058, Mar. 1953.
- [29] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, “On the optimality of the binary reflected Gray code,” *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.
- [30] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, “Gray coding for multilevel constellations in Gaussian noise,” *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 224–235, Jan. 2007.
- [31] A. Guillén i Fàbregas and A. Martinez, “Derivative of BICM mutual information,” *IET Electronics Letters*, vol. 43, no. 22, pp. 1219–1220, Oct. 2007.
- [32] M. Ivanov, F. Brännström, A. Alvarado, and E. Agrell, “General BER expression for one-dimensional constellations,” in *IEEE Global Telecommunications Conference (GLOBECOM)*, Anaheim, CA, Dec. 2012, available at <http://arxiv.org/abs/1210.8326>.
- [33] A. Lapidoth, *A Foundation in Digital Communication*, 1st ed. Cambridge University Press, 2009.
- [34] W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill, 1987.
- [35] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, 6th ed. New York, NY: Academic Press, 2000.